

CLAIMS

What is claimed is:

1. A method comprising the steps of:

(a) converting a plurality of input audio signals into a combined audio signal and a plurality of auditory scene parameters; and

(b) embedding the auditory scene parameters into the combined audio signal to generate an embedded audio signal, such that:

a first receiver that is aware of the existence of the embedded auditory scene parameters can extract the auditory scene parameters from the embedded audio signal and apply the extracted auditory scene parameters to synthesize an auditory scene; and

a second receiver that is unaware of the existence of the embedded auditory scene parameters can process the embedded audio signal to generate an output audio signal, where the embedded auditory scene parameters are transparent to the second receiver.

2. The invention of claim 1, wherein the plurality of auditory scene parameters comprise two or more different sets of one or more auditory scene parameters, wherein each set of auditory scene parameters corresponds to a different frequency band in the combined audio signal such that the first receiver synthesizes the auditory scene by (a) dividing an input audio signal into a plurality of different frequency bands; and (b) applying the two or more different sets of one or more auditory scene parameters to two or more of the different frequency bands in the input audio signal to generate two or more synthesized audio signals of the auditory scene, wherein for each of the two or more different frequency bands, the corresponding set of one or more auditory scene parameters is applied to the input audio signal as if the input audio signal corresponded to a single audio source in the auditory scene.

3. The invention of claim 2, wherein each set of one or more auditory scene parameters corresponds to a different audio source in the auditory scene.

4. The invention of claim 2, wherein, for at least one of the sets of one or more auditory scene parameters, at least one of the auditory scene parameters corresponds to a combination of two or more different audio sources in the auditory scene that takes into account relative dominance of the two or more different audio sources in the auditory scene.

5. The invention of claim 2, wherein the two or more synthesized audio signals comprise left and right audio signals of a binaural signal corresponding to the auditory scene.

1 6. The invention of claim 2, wherein the two or more synthesized audio signal comprise three or
2 more signals of a multi-channel audio signal corresponding to the auditory scene.

1 7. The invention of claim 1, wherein the combined audio signal corresponds to a combination of two
2 or more different mono source signals, wherein the two or more different frequency bands are selected by
3 comparing magnitudes of the two or more different mono source signals, wherein, for each of the two or
4 more different frequency bands, one of the mono source signals dominates the one or more other mono
5 source signals.

1 8. The invention of claim 1, wherein the combined audio signal corresponds to a combination of left
2 and right audio signals of a binaural signal, wherein each different set of one or more auditory scene
3 parameters is generated by comparing the left and right audio signals in a corresponding frequency band.

1 9. The invention of claim 1, wherein the auditory scene parameters comprise one or more of an
2 interaural level difference, an interaural time delay, and a head-related transfer function.

1 10. The invention of claim 1, wherein step (b) comprises the step of applying a layered coding
2 technique in which stronger error protection is provided to the combined audio signal than to the auditory
3 scene parameters when generating the embedded audio signal, such that errors due to transmission over a
4 lossy channel will tend to affect the auditory scene parameters before affecting the combined audio signal
5 to improve the probability of the first receiver to process at least the combined audio signal.

1 11. The invention of claim 1, wherein step (b) comprises the step of applying a multi-descriptive
2 coding technique in which the auditory scene parameters and the combined audio signal are both divided
3 into two or more streams, wherein each stream divided from the auditory scene parameters is embedded
4 into a corresponding stream divided from the combined audio signal to form a stream of the embedded
5 audio signal, such that the two or more streams of the embedded audio signal may be transmitted over
6 two or more different channels to the first receiver, such that the first receiver is able to synthesize the
7 auditory scene using extracted auditory scene parameters having relatively coarse resolution when errors
8 result from transmission of one or more of the streams of the embedded audio signal over one or more
9 lossy channels.

1 12. A machine-readable medium, having encoded thereon program code, wherein, when the program
2 code is executed by a machine, the machine implements a method, comprising the steps of:

3 (a) converting a plurality of input audio signals into a combined audio signal and a plurality of
4 auditory scene parameters; and

5 (b) embedding the auditory scene parameters into the combined audio signal to generate an
6 embedded audio signal, such that:

7 a first receiver that is aware of the existence of the embedded auditory scene parameters can extract
8 the auditory scene parameters from the embedded audio signal and apply the extracted auditory scene
9 parameters to synthesize an auditory scene; and

10 a second receiver that is unaware of the existence of the embedded auditory scene parameters can
11 process the embedded audio signal to generate an output audio signal, where the embedded auditory
12 scene parameters are transparent to the second receiver.

1 13. An apparatus comprising:

2 (a) an encoder configured to convert a plurality of input audio signals into a combined audio signal
3 and a plurality of auditory scene parameters; and

4 (b) a merging module configured to embed the auditory scene parameters into the combined audio
5 signal to generate an embedded audio signal, such that:

6 a first receiver that is aware of the existence of the embedded auditory scene parameters can extract
7 the auditory scene parameters from the embedded audio signal and apply the extracted auditory scene
8 parameters to synthesize an auditory scene; and

9 a second receiver that is unaware of the existence of the embedded auditory scene parameters can
10 process the embedded audio signal to generate an output audio signal, where the embedded auditory
11 scene parameters are transparent to the second receiver.

1 14. A method for synthesizing an auditory scene, comprising the steps of:

2 (a) receiving an embedded audio signal comprising a combined audio signal embedded with a
3 plurality of auditory scene parameters, wherein a receiver that is unaware of the existence of the
4 embedded auditory scene parameters can process the embedded audio signal to generate an output audio
5 signal, where the embedded auditory scene parameters are transparent to the receiver;

6 (b) extracting the auditory scene parameters from the embedded audio signal; and

7 (c) applying the extracted auditory scene parameters to the combined audio signal to synthesize an
8 auditory scene.

9 15. The invention of claim 14, wherein the plurality of auditory scene parameters comprise two or
10 more different sets of one or more auditory scene parameters, wherein each set of auditory scene

parameters corresponds to a different frequency band in the combined audio signal such that the auditory scene is synthesized by (1) dividing the combined audio signal into a plurality of different frequency bands; and (2) applying the two or more different sets of one or more auditory scene parameters to two or more of the different frequency bands in the combined audio signal to generate two or more synthesized audio signals of the auditory scene, wherein for each of the two or more different frequency bands, the corresponding set of one or more auditory scene parameters is applied to the combined audio signal as if the combined audio signal corresponded to a single audio source in the auditory scene.

16. The invention of claim 15, wherein each set of one or more auditory scene parameters corresponds to a different audio source in the auditory scene.

17. The invention of claim 15, wherein, for at least one of the sets of one or more auditory scene parameters, at least one of the auditory scene parameters corresponds to a combination of two or more different audio sources in the auditory scene that takes into account relative dominance of the two or more different audio sources in the auditory scene.

18. The invention of claim 15, wherein the two or more synthesized audio signals comprise left and right audio signals of a binaural signal corresponding to the auditory scene.

19. The invention of claim 15, wherein the two or more synthesized audio signal comprise three or more signals of a multi-channel audio signal corresponding to the auditory scene.

20. The invention of claim 14, wherein the combined audio signal corresponds to a combination of two or more different mono source signals, wherein the two or more different frequency bands are selected by comparing magnitudes of the two or more different mono source signals, wherein, for each of the two or more different frequency bands, one of the mono source signals dominates the one or more other mono source signals.

21. The invention of claim 14, wherein the combined audio signal corresponds to a combination of left and right audio signals of a binaural signal, wherein each different set of one or more auditory scene parameters is generated by comparing the left and right audio signals in a corresponding frequency band.

22. The invention of claim 14, wherein the auditory scene parameters comprise one or more of an interaural level difference, an interaural time delay, and a head-related transfer function.

1 23. The invention of claim 14, wherein the embedded audio signal was generated by applying a
2 layered coding technique in which stronger error protection was provided to the combined audio signal
3 than to the auditory scene parameters, such that errors due to transmission over a lossy channel will tend
4 to affect the auditory scene parameters before affecting the combined audio signal to improve the
5 probability of a receiver to process at least the combined audio signal.

1 24. The invention of claim 14, wherein the embedded audio signal was generated by applying a
2 multi-descriptive coding technique in which the auditory scene parameters and the combined audio signal
3 were both divided into two or more streams, wherein each stream divided from the auditory scene
4 parameters was embedded into a corresponding stream divided from the combined audio signal to form a
5 stream of the embedded audio signal, such that the two or more streams of the embedded audio signal
6 may be transmitted over two or more different channels to a receiver, such that the receiver is able to
7 synthesize the auditory scene using extracted auditory scene parameters having relatively coarse
8 resolution when errors result from transmission of one or more of the streams of the embedded audio
9 signal over one or more lossy channels.

1 25. A machine-readable medium, having encoded thereon program code, wherein, when the program
2 code is executed by a machine, the machine implements a method for synthesizing an auditory scene,
3 comprising the steps of:

4 (a) receiving an embedded audio signal comprising a combined audio signal embedded with a
5 plurality of auditory scene parameters, wherein a receiver that is unaware of the existence of the
6 embedded auditory scene parameters can process the embedded audio signal to generate an output audio
7 signal, where the embedded auditory scene parameters are transparent to the receiver;

8 (b) extracting the auditory scene parameters from the embedded audio signal; and

9 (c) applying the extracted auditory scene parameters to the combined audio signal to synthesize an
10 auditory scene.

1 26. An apparatus for synthesizing an auditory scene, comprising:

2 (a) a dividing module configured to (1) receive an embedded audio signal comprising a combined
3 audio signal embedded with a plurality of auditory scene parameters, wherein a receiver that is unaware
4 of the existence of the embedded auditory scene parameters can process the embedded audio signal to
5 generate an output audio signal, where the embedded auditory scene parameters are transparent to the
6 receiver and (2) extract the auditory scene parameters from the embedded audio signal; and

- 7 (b) a decoder configure to apply the extracted auditory scene parameters to the combined audio signal
8 to synthesize an auditory scene.

CONFIDENTIAL